Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 10 (2012) 1221 – 1227

Procedia Computer Science

The 1st International Workshop on Agent-based Mobility, Traffic and Transportation Models, Methodologies and Applications

An Agent-Based Model to Evaluate Carpooling at Large Manufacturing Plants

Tom Bellemans[a], Sebastian Bothe[b], Sungjin Cho[a], Fosca Giannotti[d], Davy Janssens[a], Luk Knapen[a,*], Christine Körner[b], Michael May[b], Mirco Nanni[d], Dino Pedreschi[c], Hendrik Stange[b], Roberto Trasarti[d], Ansar-Ul-Haque Yasar[a], Geert Wets[a]

[a]*Traffic Research Institute IMOB, Hasselt University, Wetenschapspark 5 bus 6, B3950 Diepenbeek (Belgium)*
[b]*Fraunhofer IAIS, Schloss Birlinghoven, 53754 Sankt Augustin (Germany)*
[c]*Dipartimento di Informatica KDD-Lab, Universit di Pisa (Italy)*
[d]*ISTI-CNR Istituto di Scienza e Tecnologie dell'Informazione A. Faedo Pisa (Italy)*

Abstract

Carpooling is thought to be part of the solution to resolve traffic congestion in regions where large companies dominate the traffic situation because coordination and matching between commuters is more likely to be feasible in cases where most people work for a single employer. Moreover, carpooling is not very popular for commuting. In order for carpooling to be successful, an online service for matching commuter profiles is indispensable due to the large community involved. Such service is necessary but not sufficient because carpooling requires rerouting and activity rescheduling along with candidate matching. We advise to introduce services of this kind using a two step process: (1) an agent-based simulation is used to investigate opportunities and inhibitors and (2) online matching is made available. This paper describes the challenges to build the model and in particular investigates possibilities to derive the data required for commuter behavior modeling from big data (such as GSM, GPS and/or Bluetooth).

Keywords:
agent-based modeling, carpooling, big data, optimization, profile matching

1. Introduction

Regions with large companies dominating the traffic situation often suffer from traffic congestion. For example, Volkswagen (VW) headquarters are located in the city of Wolfsburg, a relatively small German city with 120,000 inhabitants. The VW factory, the largest car factory in Europe, has around 50,000 workers

☆The research leading tot these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 270833

*Corresponding author: Tel.:+32-11-269-126; Fax: +32-11-269-199
E-mail: luk.knapen@uhasselt.be

1877-0509 © 2012 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.procs.2012.08.001

and employees. As a consequence, there are huge traffic problems in that area every day, and traffic patterns in the area are fully dominated by commuting patterns of VW employees. VW and the city of Wolfsburg are currently looking for solutions to these problems [1].

A similar situation exists in the city of Ingolstadt (125,000 inhabitants), where Audi employs around 30,000 persons [2], and in Genk (Belgium), where Ford employs 5000 people [3]. The problem can be generalized to many large manufacturing plants with thousands of workers in shift work, which are located in or near smaller cities. It also exists for large cities such as Cologne, but the effects are more pronounced if the factory workers constitute a large share of the traffic in peak hours. Hence, we limit our discussion to this case.

Alleviating the traffic congestion problem in these regions will normally ask for a combination of measures specifically tailored to the region in question. In this paper we are interested in one specific measure that can offer at least a partial solution to the traffic problem: carpooling. Obviously, carpooling leads to a decrease of traffic at peak hours, and if pooling rates were high enough, it would be an economical and environmentally friendly solution that does not need large upfront investments. But in fact, currently pooling rates are sometimes even decreasing in factories compared to past times because of larger availability of cars, more flexible working habits, larger spread of the employees' homes over the region, flexibility of private life, higher anonymity among employees, etc. According to [4] only 9.4% of the respondents participated in carpooling and 64% of those (5% of the population) did so at least once a week.

A promising possibility to get higher acceptance for carpooling is automatic profile construction and matching among employees that (in a 30,000-50,000 employee factory) often do not know each other personally. At the same time, because a congestion in the city of Wolfsburg or Ingolstadt with higher probability is a congestion caused by employees working for a single company, there is also a high probability of finding matching profiles. Moreover, the closed world of a company makes it much easier to increase the share of the traffic participants in pooling. Furthermore, the company might have some possibilities in experimenting with different work schedules etc. to investigate the impact of such decisions on the traffic situation. The sheer size of the problem makes automatic profiling and matching an inevitable solution. In order to solve this problem, we propose to make use of concepts from the agent-based paradigm.

The *agent-based* paradigm allows for interaction between intelligent entities which can be individuals, companies, services etc, each one having its own *belief* (perception of the environment), *desire* (ultimate goals) and *intentions* (short term tactics to achieve the goals). A behavioral model can be specified for each agent as well as the protocols to negotiate and cooperate ([5],[6],[7]). Finally, *activity based* modeling using the *agent-based* paradigm allows for modeling interactions between individuals which in turn allows the study of carpooling inhibiting factors.

Implementing large scale agent-based simulators is a challenging task due to both the computational complexity and the data requirements. In this project we plan to extract data to feed the behavioral model from available big data sets that can be recorded automatically because those sources are less expensive and less error prone than interactively conducted surveys. The project described here, is work in progress: currently the conceptual modeling phase is being excuted.

This paper analyses the factors influencing the choice for carpooling and describes an agent-based simulation model to analyze them before building an online profile matching facility. After describing the state-of-the-art in Section 2, we describe our agent-based simulation model in Section 3. In Section 4 we argue that the new data sources such as cell phone (GSM), GPS and Bluetooth data allow for much more detailed modeling of the study area and thus facilitate solutions that have not been available before due to lack of sufficient empirical information. We conclude our paper with a short summary.

2. State-of-the-art

Many public web based systems for carpooling candidate matching do exist all over the world: examples are [8] and [9] that also focuses on companies. Some of those already operational systems provide closed community services for people working for the same employer. [10] offers such services to commuters in partnership with employers, institutions, and regional governments. Partners pay a yearly fee and have access to their data to generate reports. The service is free for commuters. Furthermore, many carpooling communities and large employers offer *Guaranteed Ride Home* (*GRH*) solutions to overcome unexpected events that would prevent the return trip to take place. [11] describes *GRH* programs as inexpensive insurance policies: rules to participate, *GRH* use and associated costs are explained.

On the other hand, research focuses on inhibiting factors ([12],[13]) and feasibility ([14]). Kamar and Horvitz [15] investigate computational methods for guiding collaboration on shared plans in real-world settings where agents have diverse and varying goals preferences and availabilities. The ideas are applied to carpooling. Supply and demand for rides are posted to a system that tries to find the optimal grouping of self-interested people based on travel cost, time cost and cognitive cost. Efficiency of the resulting mechanism is calculated in terms of total cost and in terms of trip number reduction.

Public and company specific solutions try to cover carpooling inhibiting factors such as *candidate matching* and the *will-i-get-back-home* fear. However, the problem of *schedule adaptation* is not covered. Carpooling requires some rerouting: first, this takes time and second a lot of trips are multi-purpose trips (bring children to school, shop, go to work). Cooperation on commuting trips requires *cooperative routing* an *cooperative scheduling*, which can turn out to be inhibiting factors. Planning systems involving multiple people can lead to notoriously difficult combinatorial problems as is known from the application field of personnel roster calculations for hospitals and schools (see [16]).

Building and deploying a carpooling candidate matching and support application is a nontrivial and costly task. Therefore, the fraction of carpoolers in the commuting society shall be estimated under several conditions. Simply providing matching facilities is not expected to increase ride-sharing drastically. Simultaneous introduction of several measures can be required but a priori results are uncertain. Moreover, agent-based modeling of commuter's behavior, work time constraints, available transportation modes and traffic network characteristics can reveal bottlenecks and indicate the weight of carpooling inhibiting factors. In cases where commuting caused by one employer dominates the regional traffic, it should be feasible to acquire most data required for simulation.

Hence a phased project is proposed for the introduction of traffic support applications in such cases. Ideas shall be evaluated using agent-based simulation in a first phase and selected proven algorithms shall be implemented in the production system. The `Janus` framework [17] has been selected to accomplish this task because of the basic infrastructure it provides, the promising benchmarks we ran, the availability of the *holon* concept and the adherence to ASPECS.The agent-based simulator serves to answer *what-if* questions about organizational aspects and infrastructure. Specific complex components like *candidate matching* and *negotiation* can be transferred directly from the simulator to the production system.

Four-step models consisting of trip *generation*, trip *distribution*, *mode selection* and *route assignment* for traffic prediction are not suited to investigate the problem at hand because they can only account for the effects of the inhibiting factors by extrapolation from historical data; hence they do not allow evaluation of specific measures. Therefore, *activity based* [18] simulation is used. Time dependent demand for each traffic mode (car, carpool, bike, walk, bus) is derived from the predicted daily agenda for each individual. This method requires a behavioral model to be specified for each individual which is a challenging task because large amounts of accurate detail data are needed.

3. Building the Carpooling Supporting Agent-based Simulation

In this section the two phases to build the carpooling simulation will be discussed.

3.1. Phase 1: Evaluation of Carpooling Solutions using What-if Scenarios

The purpose of the first project phase is to find out, using *what-if scenarios*, how the effect of inhibitors evolves as a function of *value of time* (VOT), locations of additional carpool parkings, kilometer specific cost, altered work shift times and other factors. This phase covers the *Long-term or recurring Car Pool Problem* (*LCPP*) only. Self-interesting cost aware individuals agree to ride-share on a regular basis; they need to negotiate about the route, time and mutual payments. The agent-based paradigm is needed to model the behavior and interactions between individuals. Effects can get measured in terms of overall congestion induced delay, number of vehicle-kilometers driven versus person-kilometers driven (average vehicle occupation) etc.

The simulation first generates initial *reference schedules* as an extrapolation of the past (by using marginal densities from surveys and census) using Monte Carlo microsimulation of mutually independent individuals.

After that, it evaluates carpooling under several conditions using an agent-based model of interacting individuals trying to achieve their goals while minimizing a generalized cost. For each agent, a *daily interaction network* is generated : this is the subset of the individual's social network with whom to coordinate while planning daily activities. Individuals interested in carpooling start to explore the set of carpooling candidates via the *interaction network* thereby using the *friend-of-a-friend* (FOAF) concept also. People are matched as candidate companions based on their *agent profile* consisting of socio-demographic and location related data along with explicitly stated individual preferences. The *interaction network* is built using similarity measures for both *agent socio-demo attributes* and *travel characteristics* (timed routes): hence, each agent only has a small amount of neighbours in this network which mimics reality and avoids performance problems.When matched, candidates evaluate the cost to adapt to a new route (which can take more time than the original one) and schedule (activities shifted in space-time) using *(dis-)utility* functions based on VOT and user preferences. Most important factors during this evaluation are time constraints e.g. the ones related to pick-and-drop activities or shop opening times. Individuals then offer proposals to each other without revealing all details: they only communicate preference order over several alternatives. Details about the negotiation process can be found in [7]. Apart from negotiating carpooling conditions, those agents also rate each other with respect to timeliness, safe driving, agreement compliance etc. The resulting agent reputation is used in the carpooling candidate matching process.

In order to be realistic, simulation of this kind requires a lot of data (note that locations are areas or zones, not specific addresses): *home and work locations* of employees, *personal profiles* (gender, family status, age range of the children), *employee function class attributes* (work shift times, probability for unexpected work-time changes and associated distribution for the work-time deviation from the normal one), *impedance matrices* (specifying the travel time between carpool parkings, home and work locations under different traffic conditions), probability of *multi-purpose trips* (e.g. combined pick-and-drop, shop, work) and their *points of interest* (POI) (school, shopping center), *land-use space-time data* (information about facilities in each area: school times, shop opening times, etc) and finally information about *public transportation services*.

The schedules (daily plans) resulting from simulations can be fed into a *Traffic Simulator* (*Traffic Evolver*) to evaluate the effects on the transportation system (roads, public transport services, access to carpool parkings, etc) like traffic densities on specific links.

3.2. Phase 2: Operational realtime advisor

The realtime advisor supports solving the *LCPP*: candidates specify their origin and destination locations, time constraints (in a company specific service this can be done by specifying department and work shift), potential via-stops (schools) and related additional passengers (children). They also supply a set of preference settings to be used for profile matching. Apart from proposing candidate carpool partners, the service provides support for negotiation by proposing mutual payment amounts based on the route driven and on the required schedule adaptation.

The service also provides *Daily Car Pool Problem* (*DCPP*) support in real-time. The main purpose is support for *GRH* to regular periodic *LCPP* carpoolers who got stuck unexpectedly. Maintenance of sets of backup carpool partners based on matching profiles, can ease this task. GSM or GPS data can be used to determine current availability at the right location and to consider the current traffic situation in order to make the best carpooling choices.

Finally, for both *LCPP* and *DCPP* a facility where passengers and drivers can rate each other by supplying feedback, is essential. This is used to calculate individual reputation used for profile matching.

4. Big Data Sets and Information Extraction

In order to build our agent-based simulation we have to provide detailed information about the population within our region of interest. More precisely, our goal is to build a synthetic population with similar sociodemographic, mobility and social characteristics as the real population. Each agent in our simulation represents one person of this synthetic population acting according to a given behavioral model. However, the generation of such behavioral models is challenging due to the detailed information required. To obtain such data using a classical survey is very laborious, time-consuming and expensive. Our approach is therefore to extract all necessary information from existing *big data* sources such as GSM, GPS and Bluetooth. In initial work we have analyzed traffic patterns in the greater Wolfsburg / Braunschweig region (VW), in Köln-Niehl (Ford), and in Ingolstadt (Audi) using mobile phone data, frequency counts and other data sources. This study confirmed that a combination of heterogeneous data sources (mobile phone activity, frequency modeling, sociodemographic data) provides much better insight into the traffic patterns than any single source could do. In addition, big data sources have the following advantages. First, the data collection process is passive, i.e. the observed persons are not burdened with a questionnaire. Second, the sources typically cover a large proportion of the population, providing a comprehensive picture of behavior. Finally, some of the data set are continuously collected as the information is required for another process (e.g. billing information of GSM data), which allows us to adapt the simulation model to behavioral changes over time and to evaluate the effects of the proposed car pooling solution. In the remainder of this section we will introduce different types of big data sources, discuss their strengths and weaknesses, and give an overview on existing, sophisticated methods for the extraction of semantic information from these sources for the creation of behavioral models.

Most important for our simulation are sociodemographic characteristics (e.g. gender, age, income), mobility characteristics (e.g. place of living and working, average travel distance per day, number of trips per day, preferred mode of transportation) as well as social relationships (e.g. being member of a family, bringing children to the kindergarten, having colleagues at work). Sociodemographic information is typically available for different spatial aggregation units (e.g. municipality, post code area) by national statistical authorities. In order to enrich this information with mobility and social characteristics, we will exploit GSM, GPS and Bluetooth data. Mobile phones accompany us to nearly all daily activities and are therefore convenient for the monitoring of mobility. In addition, they are widely distributed in the population, resulting in a large data sample. As the data is used for billing purposes, it may furthermore be continuously available. The spatial and temporal resolution of GSM data depends very much on the data collection method. Typically, the spatial resolution is given by the density of mobile phone cell towers and the temporal resolution by the frequency of calling. This should be sufficient to derive e.g. origin-destination (OD) matrices of commuting behavior. In addition, first studies of GSM data [19] show promising results for the extraction of semantic information as e.g. the home or work location using the frequency of calls from certain places. Nevertheless, when extracting mobility information from GSM data it has to be kept in mind that the customers of a telecommunication provider are not representative for the population, which has to be addressed during data analysis. GSM data is also very valuable to extract the network of social interactions of a user. This interaction network is especially difficult to obtain from questionnaires due to the number of implicit and explicit interactions that may occur during every single day. An approximation of this network can be obtained using the approach presented in [20]. Here the users' traces allowed to identify the spatio-temporal coincidences between them. In other words, it is possible to build a network $G = (E, N)$ where the set of nodes $N = \{n_1 \dots n_k\}$ are the users and an edge $e_{i,j}$ exists in E if exist a place where the users n_i and n_j stop at same time, i.e. they are spatio-temporally co-located.

GPS data is very precise in terms of spatial and temporal resolution and therefore allows to infer very detailed mobility information. In contrast to GSM data, it is typically collected in a mobility survey, covering only a small portion of the population over a limited period of time. In some cases, however, GPS data from telematics units which have been installed into cars, e.g. by insurance companies or for emergency help systems, is available. This data has the advantage that it is available long-term. Independent of the source of GPS data, a number of algorithms have been developed to infer semantics from the data, which are necessary in order to derive mobility characteristics for the behavioral models of the agents. In order to understand a user's mobility from raw GPS data, two tasks have to be performed: the identification of trips and the extraction of routines. A trip hereby is the movement from one location to another where a person remains over a longer period of time. In a first step, we therefore have to detect stops from the GPS data. Such an algorithm is presented in [21] with each stop denoting a sequence of GPS points which are inside a certain region for a certain amount of time. In order to annotate the trips semantically, i.e. to derive the activities that motivate a trip, [22] use an ontology containing the expert knowledge. In this way the authors are able to classify the users (e.g., commuters, tourists, etc.) and the places where they stop (e.g., home, work, etc.) by a reasoning step. Another approach has been proposed by [23]. The authors use a clustering algorithm to build a set of routines called *mobility profiles* which describe typical routes of a person.

Finally, data collected by Bluetooth can be used as big data source. It is collected by placing Bluetooth antennas in selected locations. The antennas perform periodic scans which register devices with enabled Bluetooth functionality within their range. As Bluetooth technology is provided by default in mobile phones (e.g. to connect headsets) and increasingly also in modern cars (e.g. for car entertainment systems or to provide connection to mobile phones), the technology is able to cover a considerable portion of the population. A large percentage of cars sold nowadays are already equipped with Bluetooth, for instance Ford reports that 90% of radios in new cars are Bluetooth enabled in 2011 [24]. Studies we internally run show that with the current sensors technique roughly 5% of cars driving inside a study area can be captured. Bluetooth scans provide unique device signatures and allow therefore to combine readings of several antennas to form episodic trajectories. Thus we can gather durations of stops at specific locations (e.g. the manufacturing plan or shopping locations) and the frequency of being visited by a person.

After extracting mobility and social characteristics from big data, this information will be joined with our synthetic population. However, so far we have only considered mobility characteristics of the general population. In order to build and match profiles of employees of the car manufacturing plant, we have to incorporate their specific habits, home locations and shift working hours in the synthetic population. Such data may be obtained in a collaboration with the manufacturer, possibly in aggregated form in order to preserve the privacy of employees.

Finally, we can use continuously available big data sources in order to adapt and evaluate our carpooling system. On the one hand, the system can adapt its recommendations for car pooling partners with respect the current traffic situation. On the other hand, we are able to compare the traffic situation before and after the implementation of our agent-based simulation.

5. Conclusion

In this paper we introduced an agent-based simulation model to support carpooling at large manufacturing plants. We argue that incorporating complex negotiations between agents is required for successful carpooling, because inhibiting factors like rerouting and rescheduling have to be considered. Our model relies on comprehensive information which we will obtain from big data sources as GPS, GSM and Bluetooth. These sources offer the chance to gather information for a large portion of the population. However, sophisticated extraction methods are required to derive semantic information for the agents' behavior from the raw sources. Given continuous data streams we will be able to compare the effectiveness of carpooling advices.

References

[1] Monorail: Vw-betriebsrat in berlin, `http://www.waz-online.de/.../Monorail-VW-Betriebsrat-in-Berlin` (January 2012).

[2] Ingolstadt stadt und audi arbeiten gemeinsam an lösung der verkehrsprobleme im norden, `http://www.donaukurier.de/lokales/kurzmeldungen/Ingolstadt` (July 2011).

[3] `http://www.mobiliteitsmanagement.be/ndl/toolbox/famous/fmsford.htm`.

[4] M. Cools, K. Declercq, D. Janssens, G. Wets, Onderzoek verplaatsingsgedrag vlaanderen 4.2 (2009-2010) : Tabellenrapport, Tech. rep., IMOB, Diepenbeek, Belgium (2011).

[5] J. Hackney, F. Marchal, A coupled multi-agent microsimulation of social interactions and transportation behavior, Transportation Research Part A: Policy and Practice 45 (4) (2011) 296 – 309, special Issue: Transportation and Social Interactions.

[6] M. Wooldridge, An Introduction to MultiAgent Systems, John Wiley & Sons, 2009.
[7] H. W. Chun, R. Y. Wong, N* - an agent-based negotiation algorithm for dynamic scheduling and rescheduling, Advanced Engineering Informatics (17) (2003) 1–22.
[8] Carpoolworld : Free worldwide carpool matching, `http://www.carpoolworld.com/` (2012).
[9] Djengo, carpooling solution for your company, `http://www.djengo.be/` (2012).
[10] www.carpool.ca : Formation and management of carpools, `http://www.carpool.ca/` (2012).
[11] W. Menczer, Guaranteed ride home programs, Journal of Public Transportation 10 (4) (2007) 131 150.
[12] D. Delsaer, Mobiliteitsmanagement, Bachelor : Departement Sociaal-Agogisch werk optie: Personeelswerk, XIOS Hogeschool Limburg, Hasselt (2005).
[13] Vlaamse-Overheid, Traject-NV, Onderzoek naar hinderpalen en voorwaarden voor het succesvol promoten van carpooling, Tech. rep., Vlaamse Overheid (2007).
[14] R. Trasarti, F. Pinelli, M. Nanni, F. Giannotti, Mining mobility user profiles for car pooling, in: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '11, ACM, New York, NY, USA, 2011, p. 11901198.
[15] E. Kamar, E. Horvitz, Collaboration and shared plans in the open world: Studies of ridesharing, in: Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence, 2009.
[16] E. Ozcan, Memetic algorithms for nurse rostering, in: LECTURE NOTES IN COMPUTER SCIENCE. SPRINGER-VERLAG, THE 20TH ISCIS, LNCS (Lecture Notes in Computer Science), Istanbul, 2005, pp. 482–492.
[17] M. Cossentino, N. Gaud, S. Galland, A. Koukam, ASPECS: an agent-oriented software process for engineering complex systems how to design agent societies under a holonic perspective, Auton Agent Multi-Agent Syst (Springer Science+Business Media) 20 (2010) 260304. `doi:10.1007/s10458-009-9099-4`.
[18] Y. Shiftan, G. D. Jong, R. Europe, M. Ben-akiva, Activity-Based modeling as a tool for better understanding travel behaviour Activity-Based modeling as a tool for better understanding, Behaviour (August) (2003) 1015.
[19] C. Song, Z. Qu, N. Blumm, A.-L. Barabási, Limits of Predictability in Human Mobility, Science 327 (5968) (2010) 1018–1021.
[20] D. Wang, D. Pedreschi, C. Song, F. Giannotti, A.-L. Barabási, Human mobility, social ties, and link prediction, in: Apté et al. [25], pp. 1100–1108.
[21] V. Bogorny, B. Kuijpers, L. O. Alvares, St-dmql: A semantic trajectory data mining query language, International Journal of Geographical Information Science 23 (10) (2009) 1245–1276.
[22] M. Baglioni, J. A. F. de Macêdo, C. Renso, R. Trasarti, M. Wachowicz, Towards semantic interpretation of movement behavior, in: M. Sester, L. Bernard, V. Paelke (Eds.), AGILE Conf., Lecture Notes in Geoinformation and Cartography, Springer, 2009, pp. 271–288.
[23] R. Trasarti, F. Pinelli, M. Nanni, F. Giannotti, Mining mobility user profiles for car pooling, in: Apté et al. [25], pp. 1190–1198.
[24] `http://www.pressrelations.de/new/standard/result_main.cfm?aktion=jour_pm&r=460178`.
[25] C. Apté, J. Ghosh, P. Smyth (Eds.), Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, 2011, ACM, 2011.